The Relationship between Self-, Peer-, and Teacher-assessments in the Speaking of EFL Learners

Maryam Zakian

Abbas Moradan

Seyed Esmaeel Naghibi

Abstract

Despite the importance and broad studies on tests and assessments during so many years, educationalists, including language learners have a negative attitude towards the final teacher- administered tests. This study is conducted to investigate assessment in ELT classes for the purpose of satisfying language learners with their marks through comparing and contrasting the marks given by three groups of assessors (self-, peer-, and teacher-assessments). This has been done with 25 upper-intermediate level students at Semnan University. A questionnaire was used to compare and contrast the learners' attitudes towards their involvement in assessment. Since they were neither knowledgeable enough to assess themselves and their peers, nor aware of the experts' criteria for assessing speaking, the learner-assessors had a training session before the assessment task. The findings show that, when assessment criteria are firmly set and students are trained enough to do self and peer assessments, a strong correlation between self, peer, and teacher-assessments can be estimated through Pearson Product Moment Correlation Coefficient formula. The final mark is a kind of a mutual agreement between the teacher and the student. Self and peer-assessments would also increase the students' reflective capacity about their own learning. Involving students in the assessment makes the testing environment safe and stress free. Replacing the student-centered assessment with teacher-centered assessment removes the assessment burden. So, the marking load is decreased and the teacher finds more time to improve his/her teaching.

Keywords: self-assessment, peer-assessment, teacher-assessment, rater training, correlation coefficient, speaking

1. Introduction

1.1. The importance of assessment

The importance of assessment is clear for everyone, especially those who deal with teaching and learning in both educational and non educational settings. A mom who tries to teach her child how to say "DADY" is, in fact, trying to check the baby's progress in producing that simultaneously. First of all, she produces, repeats, and practices the word for several times, and then she asks the child to say it in various contexts through different ways, like giving a toy as a reward and showing the intended object or person to elicit the target word. Therefore, learning and assessments start from the beginning, from birthday. When it comes to the educational contexts, the same happens to the child; in

these contexts, learning and testing become more systematic and follow special rules and disciplines. The child goes to school for the purpose of learning; the teachers as well as the parents rate his/her progress occasionally. If the learning happened, and the outcomes were satisfactory, they go for further teaching, otherwise, they make some specific decisions for solving the problem and removing the learning obstacles. The importance and reasons of assessment are stated by some authorities in the field. For example, Brinke et al., (2007) cite: "Assessments are at the core of the educational process because they have a direct impact on the learning processes of students". It is also emphasized by Dahlgren (1984): "Assessment of students' learning is an important task. It has a strong influence on students' approaches to learning and outcomes of their studies".

The impact of assessment on the future of learning is also pointed by Cullingford (1997):

> "By our informal day–to–day judgments as well as by systematic evaluation processes in education, and by the ever present process of assessment, we are to a considerable extent determining the future. We need to treat that process seriously".

Coombe et al.,(2007, p. xiii) believe that the term assessment brings the image of an end–of–course paper–and-pencil tests to mind which aims at telling both teachers and students how much the student knows and does not know. They add that assessment is more than test, and it includes a range of activities and tasks that are used by teachers to evaluate a student progress and growth on a daily basis.

1.2. Methods of assessment

A vast number of studies have been conducted on the issue of assessment. For instance, Brown (2004), Bachman and Palmer (1989), Falchikov and Boud (1989), Coombe et al., (2007), Harris (2007), Weigle (2007) and many others have written books and articles about theories, types, purposes, functions and uses of assessment and test in ELT. The applications of these theories and findings have been observed in educational centers for many years.

Some types and functions of assessment have been summarized as in the following:

It has been distinguished between summative and formative evaluations, the former as the measure of what a student has grasped at the end of the unit of instruction and the latter as the measure of evaluating students in the process of forming their skills and competencies (learning).

Another distinction is made between formal and informal assessment. Informal assessment is when a teacher assesses the performance through exercise or test and helps the learner remedy the problem. The result is negotiated only to the student. But formal assessment occurs in some official ways with standardized administration procedures at the middle or the end of the term.

The next classification is that of group administered (given to the entire class at once) or individually administered test (given to one student at a time). The other classification would be: the process-based versus the product-based assessment. The former happens during the process of learning and development of the performance, and the latter happens at the end of the learning process and the focus would be on the final product rather than what went on during the whole course of study.

Farhady et al., (1994) have made another classification for the purpose of assessment:

- Achievement tests which measure the degree of students' learning from a particular instructional materials.

- Proficiency tests are used to measure the overall language ability of the learners.

- And knowledge tests which are used in situations where the medium of instruction is a language other than the learners' mother tongue.

As noted in the literature, the achievements can be scored objectively or subjectively. In objectively scored tests the scorers' opinions do not interfere with the results of the scoring, but in subjective scoring the assessors decide for the score based on his/her decision or opinion.

The type of assessment in this study based on the aforementioned classifications has been product-based and individually administered test. In terms of being formal or informal and summative or formative, it should be stated that in this study one could not draw a border line between the two since the nature of oral performance and essay writing performance is somehow different from other types of performances; therefore, the assessment of these performances would be different, too. i.e. the assessment of these performances would occur at the end of the term, but since it is really time taking to administer the test individually it has been started from the middle of the term. As a result, one can claim that it is formative and informal. At the same time, since it is related to the final product of the performance the learners' ability would not change dramatically from the middle to the end of the term, it can be claimed that it is summative and formal.   In addition, in terms of being objective or subjective in scoring performances, it should be noted that in this case the performances are assessed semi-objectively. Since there can be some degree of personal opinion interference in the assessment it is not totally objective.

The purpose of self-, peer-, and teacher–assessments in this study was to grade learners and to record marks for the purpose of comparing the scores with each other. So the main focus was on the testing of the performance rather than the improvement of the learning. This study focused on assessment of learning compared with assessment for learning, two terms distinguished by Wikstorm (2007) as:

> "Assessment of learning: assessment where not all elements of assessment process are presented to a student, assessment serves as a purpose of giving grades or satisfying the accountability demands of an external authority and is given at a specific time. Assessment for learning on the other hand is: assessment where elements of the assessment process are shared with a student, assessment serves to improve students' understanding, presenting a part of instruction that is supporting learning and is ongoing".

Studies show that the end of course teacher-assessment provokes a sense of nervousness for the learners. For example, Brown (2004), states that there is a sense of anxiety among learners about the assessment and test. In order to remove the negative sense from the learners they should be involved in the assessment process. They should carry the responsibility of the assessment themselves and do their teachers' job of assessment. (pp.1-2)

Brown (2005), Chen (2008), Matsuno (2009), Janssen–VanDieten (1989) and many others have conducted similar studies. In any of these studies, the researchers focus on some specific aspects. For instance, Patri (2002) focused on the influence of peer feedback on self– and peer–assessments of oral skills. Janssen–VanDieten (1989) focused on the validity of self–assessment in his study. Saito (2008) worked on the training and its influence on peer–assessment. AlFallay (2004) studied about the role of some psychological and personality traits of the rater on the accuracy of self– and

peer–assessments, as an instance. These studies have been done with EAP (English for Academic Purposes) and ESP (English for Specific Purposes) students mostly rather than English language learners.

Additionally, none of the studies mentioned above have been conducted in the academic contexts in Iran. Iranian students are still following the traditional educational system. In this system, the students are responsible for learning and teachers are responsible for teaching and testing. It seems that in order to make the assessment student–centered, a change of orientation is needed.

1.3. The theoretical background

Reviewing the history of assessment in the last decade, it is apparent that the concept of assessment has been changed with the change in educational practice. It has been stated to move away from a teacher–centered perspective on education to one that involves more peer interaction and group activities.

Finch, A. (2007) considers the traditional view of assessment as external and "teacher–centered" in which the teacher or external examiner designs and administers the tests. He advocates the student–centeredness and autonomy as the modern views of assessment in which students need to be involved in assessment at every point.

New developments in education have emphasized student–centered learning over teacher–centered learning. Sablonniere et al., (2009) refer to teacher–centered and student–centered approaches as the two hotly debated philosophies of education. In the former, the teacher is considered as the main source of knowledge, and the learner as a recipient only. In contrast, the latter focuses mainly on the needs and abilities of the learners, considering the teacher as the facilitator and guide in the process of learning. In student-centered approach, the students are involved in the evaluation. They are allowed to scrutinize themselves and adjust their performances.

This study is also related to cooperative or collaborative assessment. Wikstorm (2007) holds that assessment becomes "more collaborative enterprise": that is, the student is encouraged to have some degree of involvement in the process of assessment.

As another form or name for the modern type of assessment, alternative assessment presents the form of assessment that requires students to produce instead of selecting from a list of responses. Alternative assessment asks students to show what they can do, students are evaluated on what they integrate and produce rather than on what they are able to recall.

1.4. The statement of the problem

Despite all of the previous research findings and applications, there is a sense of disapproving about assessment for students. One reason can be the weakness of the tests designed to measure the learner ability. The low validity and reliability of such tests may be the cause of inaccurate estimation of learner ability.

The next can be the inability of the raters in testing. Most of the language learners in their language classes are not satisfied with the final marks given by their teachers; they think they deserve more. This might be according to the teachers' inability to rate the true ability of the learners (or the true score). Therefore, the students claim their knowledge is more than what is estimated.

The last reason can be the revolution in viewpoints towards the roles of the teachers and

learners in educational settings. Contrary to the traditional teacher–centered teaching and assessment, the educational system is the witness of the more modern and student–centered type of assessment. This type of education leads to learner autonomy as is the case in most European countries.

This study focuses on the last two causes of learner dissatisfaction. If you ask your students to choose between these dichotomies, which of them would be more favorable to them in your opinion? (Traditional vs. modern assessment, teacher–centeredness vs. student-centeredness, subjective vs. objective measurements)

Based on the researchers' previous surveys (Roberts, 2006; Cambra-Fierro, 2007; Birdsong and Sharplin, 1986), the students are inclined more towards the second part of each dichotomy. The learners do not like their scores on the final exams mainly; most of them do not trust their teachers in marking their performances. As a result, they are afraid of the exam sessions; they think even though they try hard, they will not receive a satisfactory mark. Since attending the exam session is a must to pass the university courses, the learners or let's better say, the assessees go to the exam with a sense of anxiety. This happens to the majority of the students but not all of them.

It seems that the university students are tired of the traditional roles of the teachers and students. In traditional educational settings, the teacher is the transmitter of knowledge and the learners are the passive recipients of that knowledge. In the exam sessions, the teacher plays the role of assessor, and every individual student should provide the teacher what she/he has memorized during the term either rotely or meaningfully. This situation is really boring to both learners and teachers.

The students even do not know how the teacher evaluates their performances. Have you ever heard some students say: "I am sure that the professor hates me but loves other students? So, he gave me a lower mark although I tried and studied more".

This and other similar statements show that students are not satisfied with the marks they receive on the final exams. They suppose that the teachers do not have clear and objective criteria to do the assessment job. Therefore, it seems unfair to learners. In such a situation, what would be a good solution to the problem? In other words, what should the teachers do in order to avoid such problems and keep learners or assessees satisfied?

2. Method

2.1. Research questions and hypothesis

In developing the study, these questions have been the main concerns:

1. Is there any relationship between self- and peer-assessments, self- and teacher-assessments, and peer- and teacher-assessments in speaking skill?

2. Do the learners have positive attitudes towards self-, peer-, and teacher-assessments?

The following hypotheses were devised based on the aforementioned questions:

1. There is no relationship between self- and peer-assessments, self- and teacher-assessments, and peer- and teacher-assessments in speaking skill.

2. The learners have positive attitudes only towards self- and peer-assessments.

2.2. The participants

The study had been conducted with a group of mixed 25 EFL undergraduate students studying literature at the department of foreign languages at Semnan University.

The participants were supposed to pass the course of speaking. They were homogeneous in terms of their levels of proficiency. Although the participants of the study were at upper-intermediate level, they were grouped into three levels of week, strong, and medium. In other words, they were homogeneous among which there were strong to weak students.

2.3. The instruments

A questionnaire was utilized for the purpose of comparing and contrasting the participants' attitudes towards the integration of self-, peer-, and teacher-assessments. The validity of the questionnaire had been obtained before distributing among the participants by giving it to several experts in the field.

An oral assessment rating scale (Brown and Baily, 1984, cited in Brown, 2004, pp. 142-173) and a grading chart were utilized for the purpose of judging the performance and recording the scores of each lecturer.

2.4. Procedure

The first step in the process of the study was the distribution of the questionnaire within the 15-minute time limit. After gathering the questionnaire, the overall procedure was explained to the participants. Then, the assessment criteria were explained to and discussed with the students in class with reference to the assessment forms (see appendixes A, B). Since the learners were from the same language background the scale and the instruction on how to use it were presented in the native language. The training for the assessment procedure was provided to ensure that students understood what and how to assess. In so doing, students were introduced to the criteria and were asked to practice assessment before undertaking the assessment task. All assessors were asked to use the same evaluation form to record analytical scores and comments.

Having explained the way of using the chart and scale, the participants were asked to score the performances out of 25 which would be ultimately converted to 20 as the standard top score at the academic centers in Iran. The participants were then given 10 minutes to familiarize themselves with the contents of the scale and to think and focus on the task. They were also asked to concentrate on them at home and bring their questions to the class for the purpose of clarification during next class session.

As part of the course requirement, students were assigned an oral presentation on a topic of their own interest, such as travel, education, hobbies, or club activities. They were also free to choose some topics from their text book which was "Passages 2"; they were free to choose selectively.

So, each learner was supposed to prepare and give an oral presentation lasting from 15 to 30 minutes. So, the task of oral presentation was done during the class time limit. During the assessment task, other students as well as the teacher utilized the grading chart to record the score. For the purpose of self-assessment, the voice of each lecturer was recorded and the lecturer was supposed to listen to his/her presentation and do the same as others did previously.

2.5. Data analysis

For the analysis of the data, the mean score of peer-assessment was calculated and finally,

Pearson product moment correlation coefficient formula was utilized according to the nature of the data.

3. Results

3.1. Descriptive statistics

In order to calculate the correlations between SA and PA, TA and PA, SA and TA, Pearson product moment correlation coefficient formula has been utilized through SPSS statistics 16.0. In the following tables v.1 is defined as SA, v. 2 as PA, and v. 3 as TA.

Table 1: Descriptive Statistics of 25 Students

	N	Minimum	Maximum	Mean	Std. Deviation
v.1	25	15.00	19.50	17.3760	1.48500
v.2	25	14.17	19.40	16.4712	1.25836
v.3	25	14.00	20.00	16.6000	1.85405
Valid N (listwise)	25				

v.1= SA, v.2= PA, v.3= TA

The descriptive statistics for the scores of the 25 students in speaking class has been indicated in table 1. The table provides the number of students, the minimum, maximum, mean and standard deviation of the scores in speaking group.

Table 2 indicates the correlation coefficient between SA, PA, and TA.

Table 2: Correlations between SA, PA, TA

		v.1	v.2	v.3
v.1	Pearson Correlation	1	.764(**)	.820(**)
	Sig. (2-tailed)		.000	.000
	N	25	25	25
v.2	Pearson Correlation	.764(**)	1	.791(**)
	Sig. (2-tailed)	.000		.000
	N	25	25	25
v.3	Pearson Correlation	.820(**)	.791(**)	1
	Sig. (2-tailed)	.000	.000	
	N	25	25	25

v.1= SA, v.2= PA, v.3= TA

**   Correlation is significant at the 0.01 level (2-tailed).

The above table shows that there is a significant correlation between SA, PA, and TA. The correlation between SA and PA has been reported as: .764, between SA and TA as: .820, and between TA and PA as: .791.

3.2. The analysis of the questionnaire

The analysis of the questionnaire showed that the majority of the students were not completely satisfied with teacher-given marks. Therefore, they showed positive attitude

towards self- and peer-assessments. Moreover, a large number of the respondents showed a positive attitude towards the students-given scores combined with the final teacher-assigned marks.

The majority of the respondents showed a positive attitude towards the student-given scores combined with the final teacher-assigned marks. One student claimed it is helpful in cases where teachers miss or disregard two or more points in evaluation. Some stated that it depends on the personality of the teacher. Most of the teachers make the final decisions themselves and do not like to involve students. Maybe they do so because of the students' lack of skill and experience in doing so.

A large number of the students were positive being assessed by the classmates. In the questionnaire, the respondents reported the feelings of happiness, sympathy, goodness, dislike, being honest, logical, proud, satisfied, mundane, responsible, arrogant, being helpful, serious, embarrassed, stressful and afraid of not being fair while assessed by their classmates.

More than half of the respondents reported being comfortable and making progress through self- and peer-assessments. But others felt worried, stressed, having low self-confidence, nervousness and depression. They believed self- and peer-assessments would be difficult and require experience and training.

A large number of students reported negative attitudes towards the teacher-assessment. They felt stressed and afraid of being ignored by the teacher. But an equal number stated if the teacher is a kind of person who does the assessment fairly, they feel satisfied. Additionally, if they have enough knowledge in the related course they have self-confidence and feel positive.

4. Conclusions and discussions

Despite the importance and broad studies on tests and assessments during so many years, educationalists, including language learners have a negative attitude towards the final teacher-administered tests which provoke anxiety and feelings of nervousness (Alderson, 1998; Brown, 2004).

The number of studies on the issue is vast (Matsuno, 2009; Brown, 2005; Chen, 2008; Patri, 2002). The present study sought to suggest one solution to the existing problem, and attempted to do so by involving the learners in the process of assessment and testing. The study also focused on the attitudes of the learners towards the type of assessment applied in the study.

From the results of this study which indicate a strong relationship between student-assessment and teacher-assessment besides the positive attitudes of the learners, the following conclusions and implications can be drawn.

1. Is there any relationship between SA, PA, and TA in speaking skill?

The null hypothesis for the first research question is that:

There is no relationship between SA, PA, and TA in speaking skill.

It appeared that there is a significant correlation between SA and TA, PA and TA, and PA and SA in speaking group. In speaking class, all learners were involved in assigning score to the presenter. Therefore, the peer-assessment scores were the mean sores calculated from all of the participants. There has been a range of scores given by multiple levels of students, from weak to strong ones.

The involvement of the learners in the assessment process increases the reflective capacity and the level of the critical thinking. This is parallel with what Underhill (1987, p. 24) states. He believes that "where several learners are involved in a test task at the same time they can be asked to assess each other as well as themselves. This group assessment… makes the judgment more authentic as in real life,…, using the same scale, will help the learner to make his own self-judgment more critical and accurate".

For self-assessment it should be noted that students had difficulty deciding about their own levels of ability, and tried to skip it. They were interested in doing SA for the purpose of learning rather than testing. They could not easily assign mark to their own performances.

In the process of SA there is also the danger of over- or under-estimation of students' performances. It has been observed that weak students tend to over-estimate and strong students tend to under-estimate their own abilities. These findings are confirmed by Boud and Falchikov (1989) and AlFallay (2004). This estimation has not been under the focus of this study, and concluded only subjectively. But since the scores by the teacher are very similar to the scores by learners themselves, the stated problem seems to be less likely in the current study.

The teacher-assessment in this group has been done more objectively. The teacher was asked to go through the criteria and charts as well as the learners, and apply the assessment process as accurately as possible. He tried to be more objective and excluded non-target factors that he always considered in the assessment process.

2. Do the learners have positive attitude towards self- peer-, and teacher-assessments?

The learners have positive attitude only towards self- and peer-assessments.

The analysis of the questionnaire showed that although some were not interested, most of the participants had positive attitude towards their involvement in the assessment procedure. Those who did not like it reasoned that somebody with higher ability and knowledge should judge their performances; the learners are not able to do so; they may not be fair; they lack enough confidence and experience. Also, self- and peer-assessments create stress, and friendship affects the results. But most of them were interested in doing so according to the following reasons:

1. The learners can enhance their input (in learning) by focusing on the strengths and weaknesses of their peers as well as their own.

2. It could raise the level of the students' critical thinking and reflective capacity by focusing on their own and their peers' performances and zooming on the mistakes and weaknesses.

3. This process helped students feel autonomous in the educational context. As they scored their own performances and their peers', the students felt a sense of autonomy.

4. It is better than teacher-assessment. They had less anxiety while assessed by peers than by teachers.

5. It reduced the burden of marking by the teacher. The final mark is a kind of a mutual agreement between the teacher and the student.

5. Implication for the language teachers

The high correlation between teacher and student-assessments indicates that the

teacher-assessment can be supplemented with student-assessments in the context of productive language skills.

1. If the learners can be involved in the task of assessment the teachers' time could be utilized more productively on issues related to improving their teaching techniques.

2. By involving students in the assessment process teachers can make the testing environment safe and stress free. They can involve them in the assessment of mid-term exam and benefit from the results for the final exam although they are no more involved in the final assessment.

6. Implications for the material developers

Based on the numerous studies in the realm of student-centered assessment and according to the results of the present study, some useful implications and guidelines can be stated here for material developers.

1. The material developers can include the models of assessing speaking through self and peer-assessments in the educational materials. They can develop text books or manuals concerning the importance, techniques, models, sample procedures, and training about the methods of student-centered testing and assessment.

2. They can also develop and spread the collaborative assessment in their books and gradually implement it in practice within the actual classroom assessment practice.

Since the study shows a significant correlation between self-, peer-, and teacher-assessments in speaking in cases where the instruction and training on the procedure of assessment have been applied, one can claim that the replacement of student-centered assessment with teacher-centered assessment would be a new trend in testing and assessment in the field of language teaching.

Although it seems difficult or impossible to replace the student-centered assessment with teacher-centered assessment in Iran, the educationalists can touch this new trend by integrating self-, peer-, and teacher-assessments at the early stages.

7. The last word

At this ending part of the study, it is necessary to discuss some further limitations and directions for further research that may eventually help us develop a fuller understanding of the nature of the self- and peer-assessments and their implications for language education and assessment.

1. The current study has used a relatively small sample size. So, further work needs to be carried out using a broader range of participants.

2. This study has been conducted on the participants in the upper-intermediate level; therefore, it is vital that L2 researchers devote more effort on the upper and lower levels, too.

3. The study also involves participants who lack the necessary experience in the assessment. So, it can be considered as one important factor in future studies.

References

AlFallay, I. (2004). The role of some selected psychological and personality traits of the rater in the accuracy of self- and peer-assessment. *System 32,* 407-425.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford, UK: Oxford University Press.

Birdsong, T. and Sharplin, W. (1986). Peer evaluation enhances students' critical judgment. *Highway One 9, 23-28.*

Brinke, D.J., Bruggen, J.V., Hermans, H., Burgers, J., Giesbers, B., Koper, R., & Latour, I. (2007). Modeling assessment for re-use of traditional and new types of assessment. *Computers in Human Behavior 23*, 2721–2741.

Brown, A. (2005). Self-assessment of writing in independent language learning programs: The value of annotated samples. *Assessing writing, 10*, 174-191.

Brown, H.D. (2004). *Language assessment: principles and classroom practices.* United States of America: Pearson Education,Inc.

Brown, J.D., & Baily, K.M. (1984). A categorical instrument for scoring second language writing skills. *Language learning, 34, 21-*42.

Cambra-Fierro, J. (2007). Students' self-evaluation and reflection (part 2): an empirical study. *Education and Training, 49 (2), 103-111.*

Chen, Y.-M. (2008). Learning to self-assess oral performance in English: A longitudinal case study. *Language Teaching Research,* 12, 235-262.

Coombe, C., Folse, K., & Hubly,N. (2007). *Assessing English language learners*. United State of America: University of Michigan Press.

Culling ford, C. (1997). *Assessment versus evaluation*. London: Cassell.

Dahlgren, L. (1984). *Outcomes of learning*. In The experience of learning, Edinburgh, UK (pp. 1–18). In Chang, T., and Chen, Y. (2009).Cooperative learning in E-learning: A peer assessment of student-centered using consistent fuzzy preference. *Expert systems with applications 36, 8342-8349.*

Falchikov, N., & Boud, D. (1989). Student self assessment in higher education: A meta-analysis. *Review of Educational Research, 59*, 395–430.

Farhady,H.,Jafarpur,A. , & Birjandi,p. (1994). *Testing language skills from theory to practice.* Tehran:Samt.

Finch, A. (2007). Involving language learners in assessment: A new paradigm. *English language assessment, 1 (1), 39-58.*

Janssen-van Dieten, A., (1989). The development of a test of Dutch as a second language: the validity of self-assessment by inexperienced subjects. *Language Testing 6, 30–46.*

Harris, L. (2007). Employing formative assessment in the classroom, *improving schools, 10*,249-260.

Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing,* 26, 075-100.

Patri, M. (2002). The influence of peer feedback on self and Peer-assessment of oral skills. *Language testing 19*,109-131.

Roberts, T. (2006). *Self-, peer-, and group asessment in E-learning*. United States of

America: Information science publishing.

Sablonnie` re, R., Taylor, D. M., Sadykova, N. (2009). Challenges of applying a student-centered approach to learning in the context of education in Kyrgyzstan. *International Journal of Educational Development 29, 628–63.4*

Saito, H. (2008). EFL classroom peer assessment: Training effects on rating and commenting. *Language Testing,* 25, 553- 581.

Underhill, N. (1987). *Testing spoken language: a hand book of oral testing techniques.* Britain: Cambridge university press.

Weigle,S.C. (2007). Teaching writing teachers about assessment. *Journal of second language writing, 16,* 194-209.

Wikstorm, N. (2007). "Alternative assessment in primary years of international Baccalaureate education". Master's thesis, the Stockholm Institute of Education. Thesis 15 ECTS. Online webpage citation ,http:// www.diva.portal.org.

Appendices

Appendix A: Oral proficiency scoring categories (Brown, 2001, pp. 406-407)

		Grammar	Vocabulary	Comprehension	Fluency	Pronunciation	Task
I.		Errors in grammar are frequent, but speaker can be understood by a native speaker used to dealing with foreigners attempting to speak his language.	Speaking vocabulary inadequate to express anything, but the most elementary needs.	Within the scope of his very limited language experience, can understand simple questions and statements if delivered with slowed speech, repetition, or paraphrase.	(No specific fluency description. Refer to other four language areas for implied level of fluency).	Errors in pronunciation are frequent but can be understood by a native speaker used to dealing with foreigners attempting to speak his language.	Can ask and answer questions on topics very familiar to him. Able to satisfy routine travel needs and minimum courtesy requirements. (Should be able to order a simple meal, ask for shelter or lodging, ask and give simple directions, make purchases, and tell time.)
II.		Can usually handle elementary constructions quite accurately but does not have thorough or confident control of the grammar.	Has speaking vocabulary sufficient to express himself simply with some circumlocutions.	Can get the gist of most conversations of non-technical subjects (I.e., topics that require no specialized knowledge).	Can handle with confident but not with facility most social situations, including introductions and casual conversations about current events, as well as work, family, and autobiographical information.	Accent is intelligible though often quite faulty.	Able to satisfy routine social demands and work requirements; needs help in handling any complication or difficulties.

III.	Control of grammar is good. Able to speak the language with sufficient structural accuracy to participate effectively in most formal and informal conversations on practical, social, and professional topics.	Able to speak the language with sufficient vocabulary to participate effectively in most formal and informal conversations on practical, social, and professional topics. Vocabulary is broad enough that he rarely has to grope for a word.	Comprehension is quite complete at a normal rate of speech.	Can discuss particular interests of competence with reasonable ease. Rarely has to grope for words.	Errors never interfere with understanding and rarely disturb the native speaker. Accent may be obviously foreign.	Can participate effectively in most formal and informal conversations on practical, social, and professional topics.
IV.	Able to use the language accurately on all levels normally pertinent to professional needs. Errors in grammar are quite rare.	Can understand and participate in any conversation within the range of his experience with a high degree of precision of vocabulary.	Can understand any conversation within the range of his experience.	Able to use the language fluently on all levels normally pertinent to professional needs. Can participate in any conversation within the range of this experience with a high degree of fluency.	Errors in pronunciation are quite rare.	Would rarely be taken for a native speaker but can respond appropriately even in unfamiliar situations. Can handle informal interpreting from and into language.
V.	Equivalent to that of an educated native speaker.	Speech on all levels is fully accepted by educated native speakers in all its features including breadth of vocabulary and idioms, colloquialisms, and pertinent cultural references.	Equivalent to that of an educated native speaker.	Has complete fluency in the language such that his speech is fully accepted by educated native speakers.	Equivalent to and fully accepted by educated native speakers.	Speaking proficiency equivalent to that of an educated native speaker.

Appendix B: Grading chart for analytic speaking assessment

Topic:

Date:

Rating points	Mark	Self-assessment	Peer-assessment	Teacher-assessment
Grammar	1 2 3 4 5			
Vocabulary	1 2 3 4 5			
Comprehension	1 2 3 4 5			
Fluency	1 2 3 4 5			
Pronunciation	1 2 3 4 5			
Total score out of 25	**********			

Comments: